NEW ZEALAND SCHOLARSHIP 2004

ASSESSMENT SCHEDULE FOR STATISTICS AND MODELLING

No.	Evidence	Code	Judgement
1(a)	n = 200, p = 46/200 = 0.23, z = 1.96 (95% confidence) Interval is given by: $0.23 - 1.96\sigma < \pi < 0.23 + 1.96\sigma$ Where $\sigma = \sqrt{\dfrac{0.23 \times (1 - 0.23)}{200}} = 0.0298$ So $0.17 < \pi < 0.29$ TI 83 calculator gives (0.17168, 0.28832)	BP	Confidence Interval correctly calculated – minor arithmetical error tolerated.
1(b)	Pessimistic estimate from interval is 0.172. Std Error = $\sqrt{\dfrac{0.172 \times (1 - 0.172)}{200}} = 0.027$ We require pr (proportion of support > 0.25) = pr ($z > \dfrac{0.250 - 0.172}{0.027}$) = pr (z > 2.889) = 0.5 – 0.4981 = 0.0019 The probability that the level of support is at least 25% is 0.2%. Note: π = 0.1717 gives standard error = 0.02666 so probability = 0.0017 Exact answer is 0.0022303076 Pr(x > 49.5) with continuity correction	BP	Probability correctly calculated – minor arithmetical error tolerated.
1(c)	Claim cannot be justified as 30% lies outside the upper limit of the confidence interval (0.3 > 0.29). *OR* **calculate P (p > 0.3) = 0.0094. This is very small and therefore unlikely to occur in a random sample.**	BS	Reason correctly stated as to why claim cannot be justified.

No.	Evidence	Code	Judgement
1(d)	Set prior estimate of the proportion to be 0.23 (point estimate from first survey). Set new error = 60% of 0.058 = 0.0348 Set 0.0348 = 196 x $\sqrt{\frac{0.23\times(1-0.23)}{n}}$ So $n = \left(\frac{1.96}{0.0348}\right)^2 \times 0.23 \times 0.77 = 562$ Could use another prior estimate of p = 0.29 (n = 653) with justification which gives the maximum possible sample size ***OR*** use p = 0.17 (n= 448). Note: p = 0.5 not acceptable	BP	Sample size *n* correctly calculated – minor error like one out tolerated. Variation in rounding 0.0348 tolerated in answer.
2(i)	$\frac{a}{3}+\frac{6b}{15}+\frac{7c}{18}=11.5$ Multiply through by 90 to get: 30*a* + 36*b* + 35*c* = 1035	BM	Third equation correctly deduced – can have differing multiples of the equation.
2(ii)	Set of equations reduced to two, so we have more than one solution. Solve for *b* = $\frac{279-5c}{6}$ and *a* = $\frac{-5c-639}{30}$ or *a* = - 0.17*c* – 21.30 and *b* = 46.50 – 0.83*c*	BM	General solution is correctly found.
2(iii)	Note that *a, b* and *c* all are greater than 0. The general solution for *b* implies that *c* < 55.8. Also note that *a* can never be positive from Q2 (ii) and you cannot have negative quantities. So you would have wastage of nitrogen, phosphorus or potassium to achieve compost production and/or find shortfall to make up difference by ordering more.	AC	Correctly conclude that solution is unworkable with a suggested action.

No.	Evidence	Code	Judgement
3	The following points could be covered in a one-page summary: 1. Correlation. 2. Interpret at least two distinct sections of plot. 3. Interpreting the coefficients of the regression line. 4. Summary statistics of the values of expenditures and sales. (Could include ratio of sales to expenditure) 5. Reference to outlier 6. Sales predictions. 7. Validity of sales predictions. 8. Other factors affecting sales – like seasons, market share. A complete example of a one-page summary is given below: Overall there is a moderate positive correlation between expenditure and sales. This is evidenced by an overall correlation coefficient of 0.819. In fact, there are three distinct sections of the graph with weak positive correlation for expenditures below \$15 000 and almost constant sales for expenditures in excess of \$30 000. The value of the correlation improves markedly with the removal of the outlier at (\$23 000, \$210 000). The highest correlation of 0.972 is obtained when we consider the linear scatter between expenditures of \$15 000 and \$30 000. There is a non-linear scatter where the points tail off with expenditures greater than \$30 000. The graphs suggest that there is a saturation point for sales at approximately \$180 000, which means that no matter how much more is spent on promotion, sales will not increase appreciably beyond that value. Sales prediction for say E = \$24 000 is given by: $S = 53.2 + 3.67 \times 24 = 141.28$, so predicted sales = \$141,000. This prediction would be valid as the E value lies within the linear section of the scatter that has the highest correlation coefficient. If we obtain a sales prediction for E = \$45 000 we get: \$176 000 using the scatter. This occurs in the flat part of the graph where market saturation could have occurred and may be subjected to a greater error.	BS/AT (BS if required) / AC	BS: 1, 2, 3, 4, 5 (3 out of 5) AT: 6 (At least two predictions with no more than one graphically) AC: 7, 8 – 1 out of 2.

<table>
<tr><th>No.</th><th>Evidence</th><th>Code</th><th>Judgement</th></tr>
<tr><td>3 cont.</td><td>For the equation: $S = 53.2 + 3.67\,E$, for every $1 000 increase in expenditure the model suggests sales increase by $3 670 on average. It also suggests that the base line sales with no advertising are approximately $53 000.

Summary statistics are as follows:

<table>
<tr><th></th><th>Expenditure ($000)</th><th>Sales ($000)</th></tr>
<tr><td>Mean</td><td>26.4</td><td>144.9</td></tr>
<tr><td>Median</td><td>24.0</td><td>149.5</td></tr>
<tr><td>Std Deviation</td><td>17.5</td><td>32.2</td></tr>
<tr><td>Range</td><td>0–60</td><td>98–210</td></tr>
</table>
</td><td></td><td></td></tr>
<tr><td>4(a)</td><td>Let x = number of bottles of POW and y = number of bottles of ZAP
Deduce constraints as follows:

1. $30x + 20y \le 25 \times 60$ which simplifies to $3x + 2y \le 150$
2. $x \le 35$ and $x + y \le 65$
3. $x \ge 15$ and $y \ge 15$

Profit function is $P = 20x + 10y$

So, sketch the feasible region with corner points as follows:

Going clockwise: (15,15), (35,15), (35,22.5), (20,45) and (15,50):

(a) P is maximized at (35, 22) and (34, 24); these are either side of (35, 22.5), which you cannot have due to x and y needing to be integers.
Thus the recommendation is that either 34 bottles of POW and 24 bottles of ZAP or 35 bottles of POW and 22 bottles of ZAP will maximize profit.</td><td>BM

both answer pairs required</td><td>1. Correct answers obtained with an acceptable method. Profit not required.
2. Both correct combinations specified.</td></tr>
</table>

No.	Evidence	Code	Judgement
4(b)	P becomes P = $15x + 10y$. Recommendation becomes any of the production combinations: (20, 45), (22, 42), (24, 39), (26, 36), (28, 33), (30, 30), (32, 27) and (34, 24).	AT (BM if required) – every possible combination pair identified	1. Correct answers obtained with an acceptable method. Profit not required. 2. All correct (implied integer) combinations specified in any appropriate format including line description.
4(c)	The graph $y = 9 \ln(x)$ intersects the line $3x + 2y = 150$ on the edge of the feasible region. So: $x + 6 \ln(x) - 50 = 0$	AT (BM if required)	Equation correctly deduced. Can have equivalent equations.
4(d)	To solve this equation, several methods can be employed. Some students have graphical calculators. Get 29 bottles of POW and 31 bottles of ZAP. Eg graphical technique by plotting $y = 9 \ln(x)$ on graph constructed previously plus trial by error involving substitution of values.	BM	Correct answers obtained. Allow (29, 31), (29.30) and (30, 30) as solutions. If wrong equations in (c) no credit if used to get answer. Un-rounded answers not accepted.

No.	Evidence	Code	Judgement
5(a)	1. Binomial with n = 20 and π = 0.02. So probability of truckload being accepted is $0.98^{20} + 20 \times 0.02 \times 0.98^{19} = 0.940$ Probability of one truckload being accepted = $2 \times 0.94 \times 0.06 = 0.113$ ***OR*** Use Poisson approximation to the Binomial with $\lambda = n\pi$. This gives probability of truckload as being accepted as 0.9384 and probability of one truckload being accepted as 0.116. 2. Target acceptance rate is ≥ 0.96 so use Poisson approximation with $\lambda = n\pi$. From tables, closest λ value to pr (x = 1) + pr (x = 0) being closest to 0.96 is 0.3. So π = 0.3/20 = 0.015, ie 1.5%. 3. Probability of no defectives = $0.98^{20} = 0.668$ Probability of acceptance on second check = $0.06 \times 0.668 = 0.040$. So proportion of accepted truckloads that would be accepted on the second $\text{sample} = \frac{0.040}{0.94+0.04} = 0.041$, ie P (truckload accepted from second sample/accepted) = 0.041.	1. and 2. BP in each / 3. AE (BP if required)	1. Method correctly applied. Minor variation in answer accepted 2. Method correctly applied. Minor variation in answer accepted like 1.6% or 0.015 (or equivalent) 3. Method correctly applied. Minor variation in answer accepted. Must show conditional aspect in answer
5(b)	Use Poisson approximation and look for closest pr (x = 0) + pr (x = 1) to 0.91. Observe that λ = 0.5. So $\lambda = n\ \pi = n \times \frac{1}{150} = 0.5$, so $n = 75$. So a maximum of **75 punnets** should be sampled by NAILS.	AE (BP if required)	Correct answer obtained. n=74 or n = 77 is acceptable. Variations in method accepted but must be expressed as integers.

No.	Evidence	Code	Judgement
6(a)	1. • The long-term trend, which shows a gradual increase in sales over time. • The seasonal fluctuations in the monthly retail sales, which show a high in December and lows in July/August. • The daily fluctuations over a week, like peaks in the weekend and troughs mid-week. 2. • Calculate a seasonal effect for each month and each day using the appropriate data. • Obtain a trend forecast for a particular quarter, then for the month under consideration. • 'Add' the seasonal to get a forecast for that month. • Divide to get an average daily forecast for the days in that month. • Adjust this forecast depending on the day of the week, eg if Saturday is 30% above average we multiply by 1.3. • To get a daily sales target we then multiply by 1.05.	BS	An implied comment about <u>Trend</u> plus mention all of the following at least once: <u>seasonal</u>, <u>forecast</u>, <u>adjust for day</u> and <u>inflate by 5%</u>.
6(b)	1. The high sales for Labour Day have the effect of increasing 7 values of the centred moving mean either side of the Monday, ie Friday before to Thursday after. This irregular variation in the time series would provide an inflated trend in this section of the time series for daily sales. 2. To calculate a seasonal effect for Mondays we could ignore the Labour Day sales figure and work on the other Monday sales. A trend value would be obtained for the Monday under consideration, and then the seasonal effect is 'added' into the trend to make a forecast. A special adjustment could be made for Labour Day based on previous data, ie if sales are generally 30% above the daily average for Labour Day, then multiply by 1.3.	BS	1. Key point of increasing the moving average for a week of values must be mentioned, **AND** 2. Mentions using other Mondays, then adjust specially for Labour Day. Accept an answer that doesn't specifically identify a Labour Day forecast.

<table>
<tr><th>No.</th><th>Evidence</th><th>Code</th><th>Judgement</th></tr>
<tr><td>6(c)</td><td>Either solution accepted:
Trend Value for December 2004 = 225.21 + 0.91 $\times$ 42 = 263.43.
Trend Value for each day in December
$2004 = \frac{263.43}{29} = 9.084.$
Seasonal effect for Tuesdays = -2.6 (-1.9 accepted as alternative)

Forecast for 7th December 2004 = 9.084 – 2.6 = 6.484, ie 6.5 to 1 d.p., ie $6 500.
OR

An adjustment for the month of December is included so 129.95 is added to the trend value before division into 31 days. This gives 12.69 leading to a December 7th forecast of $10,100.</td><td>BM</td><td>Correct method in either case with minor variation in answer tolerated. Accept rounding but not missing $000. Forecasts change if 30 days: $6,200 or if 31 days: $5,900. If number of months within 2 of 42 accept resulting calculations if correct. Daily trend line approach not accepted however multiplicative approach to “adding” seasonal accepted.</td></tr>
<tr><td>6(d)</td><td>1. No account is taken of the increasing trend over a particular month. It would tend to be lower than average at the start and higher at the end.
2. The gradual upwards trend is assumed to continue into late 2004.
3. The calculated seasonal effects are approximate and have been calculated over a small number of Tuesdays and/or Decembers.
4. No sales data are used between 1st May and present day, so recent sales have no effect on calculated forecast.
5. The calculated seasonal effect is unreliable due to the Labour Day increased sales.</td><td>AC</td><td>A minimum of two of these limitations required in answer. Statements where relevant must be consistent with the answer in part c.</td></tr>
</table>